

95%HPD intervals of average relative error from BEST analysis			
True inference model (Reference for misspecification experiments)	CNN APE	Posterior mean APE	CNN APE - Posterior mean APE
R_0	2.4, 3.5	2.1, 3.1	0.1, 1.2
δ	7.0, 10.5	5.7, 8.9	0.2, 3.0
m	9.5, 14.1	8.4, 12.1	0.4, 3.2
Misspecified R_0 experiment	CNN APE - CNN Reference APE	Posterior mean APE - Post. mean Reference APE	CNN APE - Posterior mean APE
R_0	11.8, 17.8	11.0, 16.9	-0.1, 1.6
δ	0.8, 7.6	-0.6, 5.3	1.3, 5.8
m	8.2, 17.9	6.5, 15.9	1.3, 4.7
Misspecified sample rate experiment	CNN APE - CNN Reference APE	Posterior mean APE - Post. mean Reference APE	CNN APE - Posterior mean APE
R_0	-0.3, 1.7	0.03, 1.7	0.1, 1.3
δ	12.0, 21.2	12.6, 21.4	0.1, 4.0
m	3.3, 12.0	5.6, 14.4	-1.2, 2.7
Misspecified migration rate experiment	CNN APE - CNN Reference APE	Posterior mean APE - Post. mean Reference APE	CNN APE - Posterior mean APE
R_0	-0.9, 0.8	-0.6, 1.0	-0.5, 0.8
δ	-2.3, 3.3	0.1, 5.8	-1.4, 2.3
m	4.0, 15.2	5.0, 16.2	-1.3, 2.6
Misspecified number of loci experiment	CNN APE - CNN Reference APE	Posterior mean APE - Post. mean Reference APE	CNN APE - Posterior mean APE
R_0	-0.3, 1.5	-0.7, 0.8	0.5, 1.9
δ	-0.3, 4.9	-0.5, 4.2	0.4, 3.5
m	3.4, 11.1	5.8, 13.5	-0.9, 1.6
Phylogenetic error experiment	CNN APE - CNN Reference APE	Posterior mean APE - Post. mean Reference APE	CNN APE - Posterior mean APE
R_0	0.7, 3.0	1.7, 4.4	-1.4, 0.1
δ	2.3, 9.6	1.5, 7.2	1.4, 5.3
m	-1.2, 6.0	-1.8, 5.4	-1.7, 2.4